Rocznik Ochrona Środowiska

Volume 26 Year 2024 ISSN 2720-7501 pp. 313-321

https://doi.org/10.54740/ros.2024.031 open access

Received: May 2024 Accepted: July 2024 Published: August 2024

Evaluation of Sulphur Dioxide Hourly Prediction Using Long Short-term Memory for Summer and Winter Season

Mohammed Bennis[1], Youssfi Mohamed[2], Rachida El Morabet[3], Majed Alsubih[4], Muneer Prayanagat[5], Roohul Abad Khan[6]*

[1]*21ACS laboratory, ENSET Mohammedia, University Hassan II Casablanca, Morocco*

[2]*21ACS laboratory, ENSET Mohammedia, University Hassan II Casablanca, Morocco*
https://orcid.org/0000-0003-2842-9880

[3]*LADES lab, FLSH-M, Hassan II University of Casablanca, Mohammedia, Morocco*
https://orcid.org/0000-0003-4750-644X

[4]*Department of Civil Engineering, King Khalid University, Abha, Saudi Arabia*
https://orcid.org/0000-0001-8226-1682

[5]*Department of Electrical Engineering, King Khalid University, Abha, Saudi Arabia*

[6]*Department of Civil Engineering, King Khalid University, Abha, Saudi Arabia*
https://orcid.org/0000-0002-2329-4123

[*]*corresponding author's e-mail: rakhan@kku.edu.sa*

Abstract: Increasing air pollution has necessitated the prediction of pollutants over time. Deterministic, statistical, and machine-learning methods have been adopted to predict and forecast pollutant levels. It aids in planning and adopting measures to overcome the adverse effects of air pollution. This study employs long short-term memory (LSTM). This study used the hourly data from a meteorological station in a low-town area, Mohammedia City, Morocco. The model prediction accuracy was evaluated based on hourly, weekly, and seasonal (summer and winter) readings for the summer and winter of 2019, 2020 and 2021. Root mean square error (RMSE), mean absolute error (MAE) and mean arctangent absolute percentage error (MAAPE) were calculated to evaluate the accuracy of the developed LSTM model. The MAE value of 0.026 was observed to be less in winter than 0.029 during summer in 2019. Also, it was observed that MAE values decreased from Year 2019-2021, indicating increased prediction accuracy. MAAPE also observed a similar trend. However, RMSE values indicated the opposite for 2019 and 2020; in 2021, the RMSE value was 0.21 for summer and 0.14 for winter for hourly readings. Based on the error calculation, the study found weekly hourly readings were the most accurate for predicting SO_2 concentration. Also, the LSTM model was more accurate in predicting winter SO_2 concentration than in the summer season. Further studies must incorporate local incidences affecting the SO_2 concentration into the LSTM model to increase its accuracy.

Keywords: sulphur dioxide, machine learning, long short-term memory, mean absolute error, root mean square error

1. Introduction

Air pollution and its adverse health effects have necessitated air quality monitoring in major urban cities (Afonso et al. 2020, Clare Heaviside et al. 2020). Sulphur dioxide is one of the recognised priority air pollutants which occurs in urban environments as it is capable of causing adverse impacts on human health and the environment (Morakinyo et al. 2020). Hence, monitoring and predicting its concentration in urban air is necessary to assess its potential health impact. Also, to aid policymakers and decision-makers in adopting policies to reduce its adverse health impact.

The forecast and prediction of SO_2 can help the population be ready to minimise the health impact of SO_2 exposures. Many deterministic methods can predict the dispersion and transport of air pollutants in the atmosphere. For example, WRF (Weather Research and Forecasting) models, NAQPMS (National Air Quality Prediction Modelling System), CTMs (Chemical Transport Models), OSPM (Operational Street Pollution Models), etc. (Ma et al. 2019). However, these prediction models application is restricted due to Operation cost (expensive), fixed parameters, i.e., pre-defined in the model and lack of accurate data (Qi et al. 2019, Zhao et al. 2019, Luo Zhang et al. 2021). Another approach was the adoption of statistical models, i.e., ARIMA (Autoregressive Integrated Moving Average), GWR (Geographically Weighted Regression), MLR (Multi Linear Regression) and GAMs (Generalized Additive Models). Most models mentioned above assume a linear relationship between the target and the variable. It is not true in real-time scenarios, which further limits the accuracy of these models (Ma et al. 2019).

To address this research gap long short-term memory neural networks have been adopted in many research works for accurate prediction and has been used successfully for prediction in several research areas; for example, Liu et al. (2021) employed LSTM for the prediction of the state of charge in lithium-ion batteries. Wang et al. (2021) employed the LSTM model for voltage stability. Zuo et al. (2021) used LSTM to predict the degradation of proton exchange membrane cells. Lei Zhang et al. (2021) predicted weather radar echo for sustainable e-agriculture by utilising LSTM networks. In air quality prediction and assessment, LSTM has also been applied; spatiotemporal air quality prediction (Seng et al. 2021), EL-nino and IOD data-based rainfall prediction (Haq et al. 2021), air quality prediction (Li et al. 2017, Ma et al. 2019, Ma et al. 2020, Zhai & Cheng 2020) bidirectional LSTM neural network for air quality prediction (Luo Zhang et al. 2021) $PM_{2.5}$ concentration prediction (Qi et al. 2019, Zhao et al. 2019) and NO_x concentration prediction (Yang et al. 2020). Also, the potential of LSTM for priority pollutant prediction is yet to be fully explored. Nevertheless, there is still a lack of literature for priority pollutants prediction using LSTM specifically for sulphur dioxide. Sulphur dioxide is used in this study for one main reason: a petroleum refinery near the prefecture, and sulphur is produced as a by-product of crude oil processing. Hence, it becomes a priority pollutant to be assessed first than other pollutants. Additionally, the LSTM application for air quality assessment has yet to be investigated in Mohammedia City, Morocco.

Hence, the novelty of this study is the application of LSTM for sulphur dioxide concentration in the urban environment of the lower town of Mohammedia City, Morocco. This study examines the performance of LSTM based on 1) hourly SO_2 concentration, 2) weekly average hourly concentration and 3) seasonal hourly concentration. The current work, being the first of its kind in Mohammedia prefecture, will serve as a reference for future studies in the region, Morocco, and especially in North African countries that experience similar climatic conditions.

2. Methods and Data

2.1. Data

The data was obtained from DGM "Direction de la météorologie nationale" Morocco in collaboration with FLSH-M. The hourly data was obtained for three years, viz. 2019, 2020 and 2021. The data was evaluated based on hourly, weekly, and seasonal (summer and winter) hourly concentrations. Fig. 1 presents the study area of Mohammedia City, Morocco. Fig. 2 presents the SO_2 concentrations during the summer and winter seasons. Hence, it can be inferred that concentration is restricted to time and varies with season.

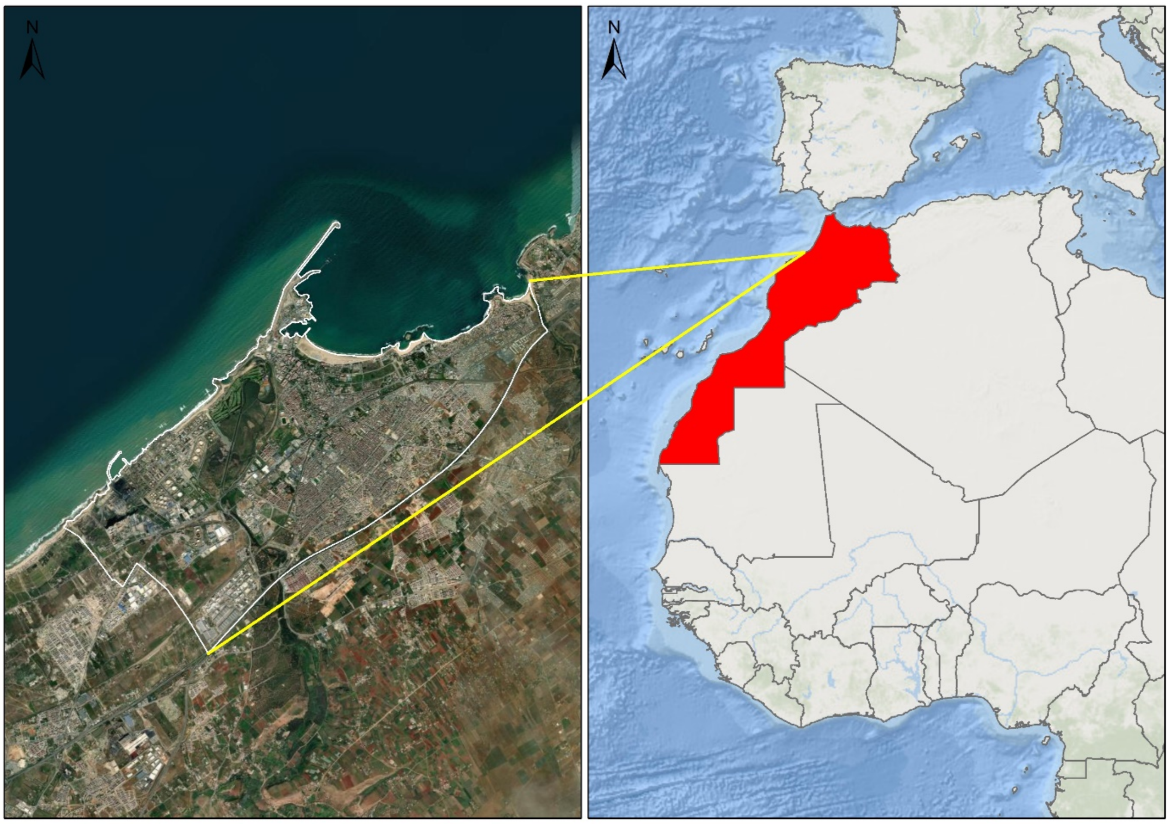
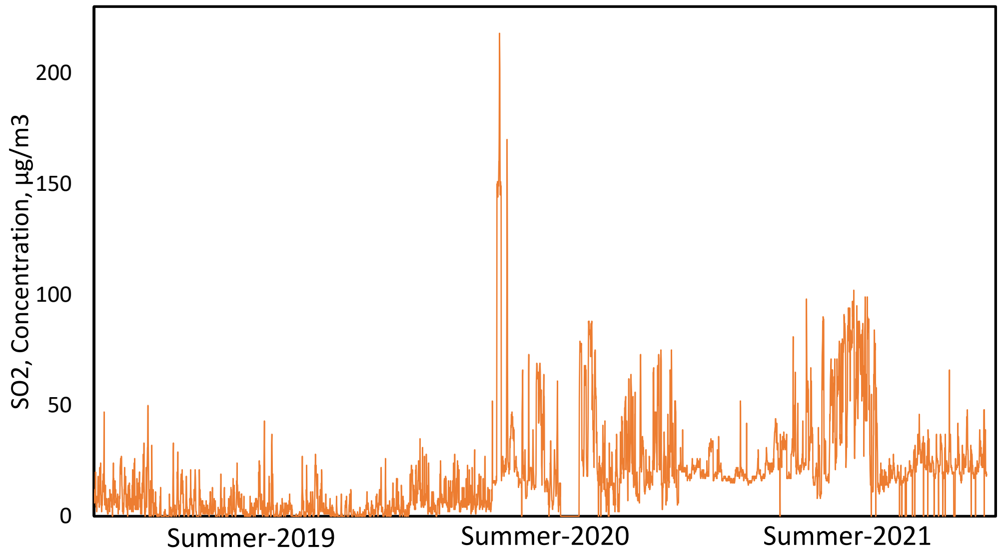


Fig. 1. Study Area map of Mohammedia City, Morocco



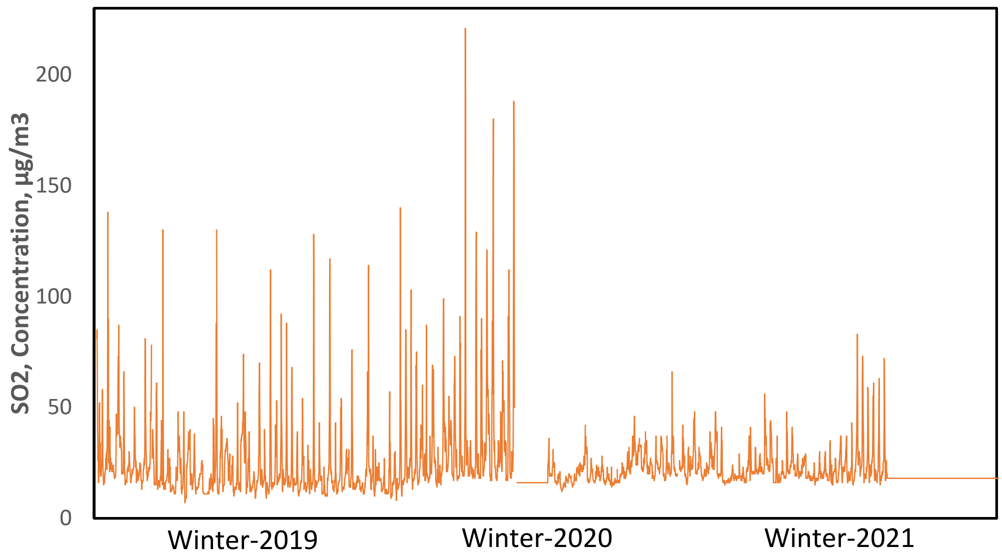


Fig. 2. Sulphur Dioxide (SO_2) concentration during summer and winter seasons for years 2019, 2020 and 2021

2.2. Long Short-term Memory (LSTM)

LSTM is a modified version of the recurrent neural network (RNN). LSTM was reliable for time series prediction and modelling based on its ability to memorise long-range dependencies (Zhao et al. 2019). The LSTM model consists of a basic unit termed a memory cell. Every memory block comprises three nonlinear gates termed the forget gate ($\mathbf{f}_t$), output gate ($\mathbf{o}_t$) and input gate ($\mathbf{i}_t$). Forget gate and input gate are controllers of unit state $\mathbf{c}_t$. While the output gate is responsible for controlling the amount of cell state $\mathbf{c}_t$, which will be mapped as $\mathbf{h}_t$ (current out value). The weight matrices are denoted by $\mathbf{W}_f$, $\mathbf{W}_i$, $\mathbf{W}_o$, and $\mathbf{W}_c$ for input vector x_t at time t. The weight matrices of hidden values are $\mathbf{U}_f$, $\mathbf{U}_i$, $\mathbf{U}_o$, and Uc, which are applied to block ht-1. Bias vectors are written as bf, bi, bo, and $\mathbf{b}_c$. The LSTM model employed in this study is the same as that employed by Qi et al. (2019) and Zhao et al. (2019) and presented in Eq. 1-5.

$$i_t = \sigma_g (W_i x_t + U_i h_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma_g (W_f x_t + U_f h_t + b_f) \tag{2}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \sigma_g (W_c x_t + U_c h_{t-1} + b_c) \tag{3}$$

$$o_t = \sigma (W_o x_t + U_o h_t + b_o) \tag{4}$$

$$h_t = o_t \cdot \tanh (c_t) \tag{5}$$

The error of predicted value was calculated by the MAAPE (mean arctangent absolute percentage error), a modified version of MAPE (mean absolute percentage error). The advantage of MAAPE is its acceptance of zero values target variable and limited value range, which eases the network training process. The estimation of MAAPE was done as per Eq. 6.

$$\text{MAAPE} = \frac{1}{N} N \sum_{i=1}^{N} arctan\left(\left|\frac{y_i - f_i}{y_i}\right|\right) \tag{6}$$

Where original values are presented as y_i, predicted data as f_i, and the number of predicted data as N. 0 to Π/2 was a range of MAAPE values. The smaller the MAAPE values, the more accurate the model (Haq et al. 2021).

Mean absolute error (MAE) and root mean square error (RMSE) were also calculated to evaluate the model performance as per Eq. 7 and 8.

$$MAE = \frac{1}{N}\sum_{k=1}^{N}|p_k - a_k| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(p_k - a_k)^2} \tag{8}$$

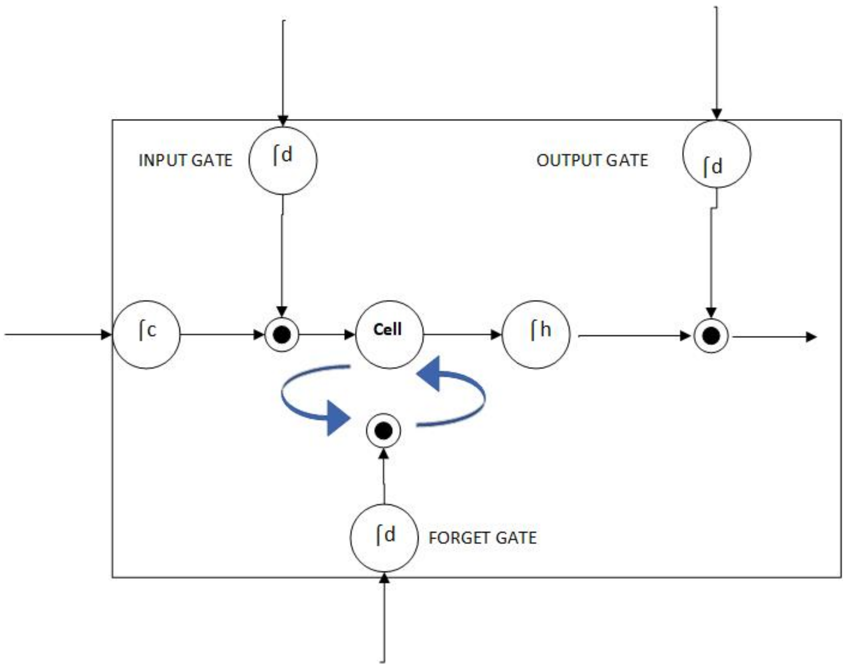


Fig. 3. Memory block of LSTM model for one cell

3. Results Discussion

3.1. LSTM model accuracy

The accuracy was further validated regarding MAAPE, MAE and RMSE values, which provide the error in prediction by the LSTM model. Table 1 shows that the least error was observed in weekly hourly concentrations followed by seasonal and hourly values. In terms of season, the LSTM performance model was more accurate for hourly, weekly, and seasonal predictions for winter than summer. Zhai & Cheng (2020) observed RMSE and MAE values of 3.53 and 2.30, respectively, for modelling SO_2 using sulphur dioxide using the LSTM model. Luo Zhang et al. (2021) have reported an RMSE value of 6.86 and an MAE value of 4.92 for modelling $PM_{2.5}$ using an LSTM semi-supervised bidirectional model. Ma et al. (2020) used a transfer learning-based stacked bidirectional LSTM model and observed that the RMSE value was reduced by 35.21%. They reported an RMSE value of 7.95 with 3 frozen layers in the LSTM model.

Table 1. LSTM accuracy based on MAE, RMSE and MAAPE values during summer and winter seasons for the year 2019, 2020 and 2021

		MAE			MAAPE			RMSE		
		2019	2020	2021	2019	2020	2021	2019	2020	2021
Summer	Hour	-0.02942	-0.00451	0.00401	0.00046	-0.00019	-0.00025	0.51273	0.22033	0.21030
	Week	0.00025	-0.00025	0.00015	-0.00003	0.00002	0.00002	0.02764	0.02319	0.00594
	Season	-0.00308	-0.00976	0.00270	-0.00034	0.00003	-0.00019	0.25609	0.24585	0.22118
Winter	Hour	-0.02668	-0.01143	0.00187	0.00034	0.00041	-0.00017	0.65099	0.31969	0.14966
	Week	-0.00176	-0.00131	0.00003	0.00003	0.00005	-0.00001	0.06459	0.05349	0.03668
	Season	-0.02708	-0.01255	0.00151	0.00030	0.00039	-0.00016	0.66154	0.36286	0.15345

3.2. Prediction performance of the LSTM model

The prediction performance of LSTM is presented in this section. The optimum prediction parameters were obtained from trial-and-error experimentation. 60% of the dataset was used for training, and 30% was used for validation. The dataset was evaluated for three consecutive years, i.e., 2019, 2020 and 2021.

Hourly prediction performance is presented in Figure 4. Hourly prediction enables visualisation of the peak of SO_2 concentration and its predicted values. The greater difference in prediction values was found at various peak hours. The greater difference was observed at midnight for the year 2019, which shifted to 4 PM – 8 PM in the year 2020 and again in the year 2021, it was found to be at noon hours. The increase in SO_2 concentration is attributed to local incidences such as traffic conditions, industrial activities, etc. At noon, the traffic volume increases, leading to increased exhaust flue gases for a longer duration than normal traffic concentration. Also, with congestion in traffic, the number of vehicles on the road increases, which means the concentration of exhaust gases also increases compared to normal traffic hours. This rise in traffic is attributed to school closing hours, lunch hours, changes in shifts or residents travelling back, and numerous activities coinciding with lung hour or break-related activities. Being at lunch hour means restaurants will be operating at their peak compared to the previous period. Also, many food outlets operate during these hours, which means high exhaust gases are attributed to outlets dealing with barbeque servings. Based on hourly concentration, it can be inferred that LSTM can predict more accurately without local incidences. This further can be validated in terms of weekly hourly concentration.

The weekly hourly concentration is presented in Figure 5. Since the weekly hourly concentration is the average of seven days for each hour, it reduces the impact of local incidences on SO_2 concentration. Hence, the performance of LSTM is substantially improved. For weekly hourly concentration, LSTM model prediction accuracy increased each year for the summer season, which can be observed from Figure 4. This can be attributed to an overall increase in SO_2 concentration, i.e., an increase in pollution, as the average value of SO_2 was 9 $\mu g/m^3$ for the year 2019, 13 $\mu g/m^3$ for the year 2020 and 18 $\mu g/m^3$ for the year 2021. This can be attributed to the increase in tourism activities experienced in the city, which greatly influences the local traffic conditions.

For the summer season, the trend was different, with the year 2019 giving better predictions than the year 2020 and 2021. The non-uniform trend of change in SO_2 concentration can be attributed to local industrial activities and the impact of local weather conditions, which greatly affect the local traffic conditions. Fig. 6 presents the SO_2 concentration against the LSTM predicted values, and it can be observed that the prediction was good for all three years and two seasons. For 2019, the prediction was relatively low as the SO_2 ambient air concentration was low, and the peak range was higher compared to other years. Hence, with a relatively lower range of change in SO_2 concentration, the prediction performance of LSTM increases.

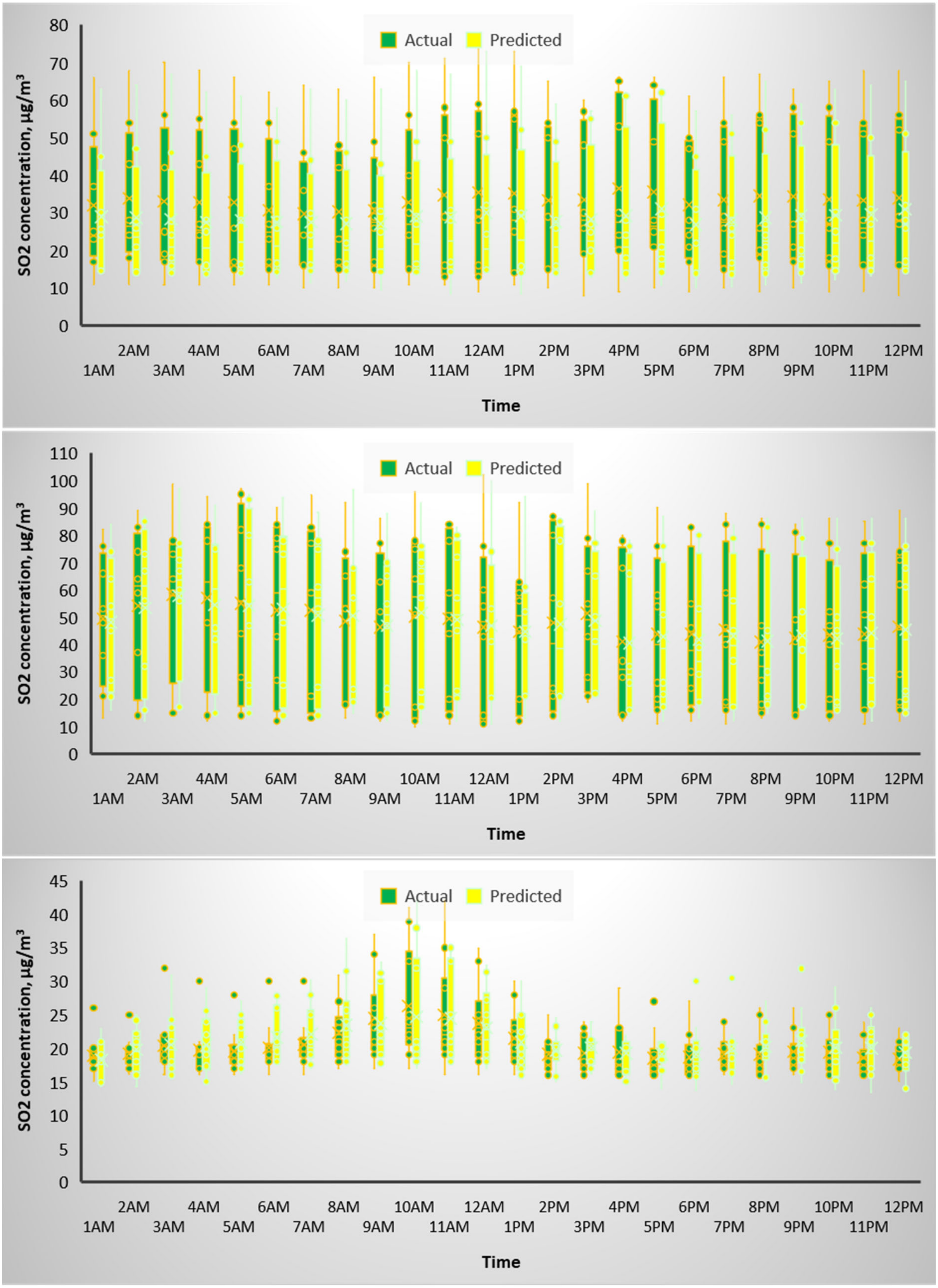


Fig. 4. SO_2 concentration actual and predicted values for years (top to bottom) 2019, 2020 and 2021

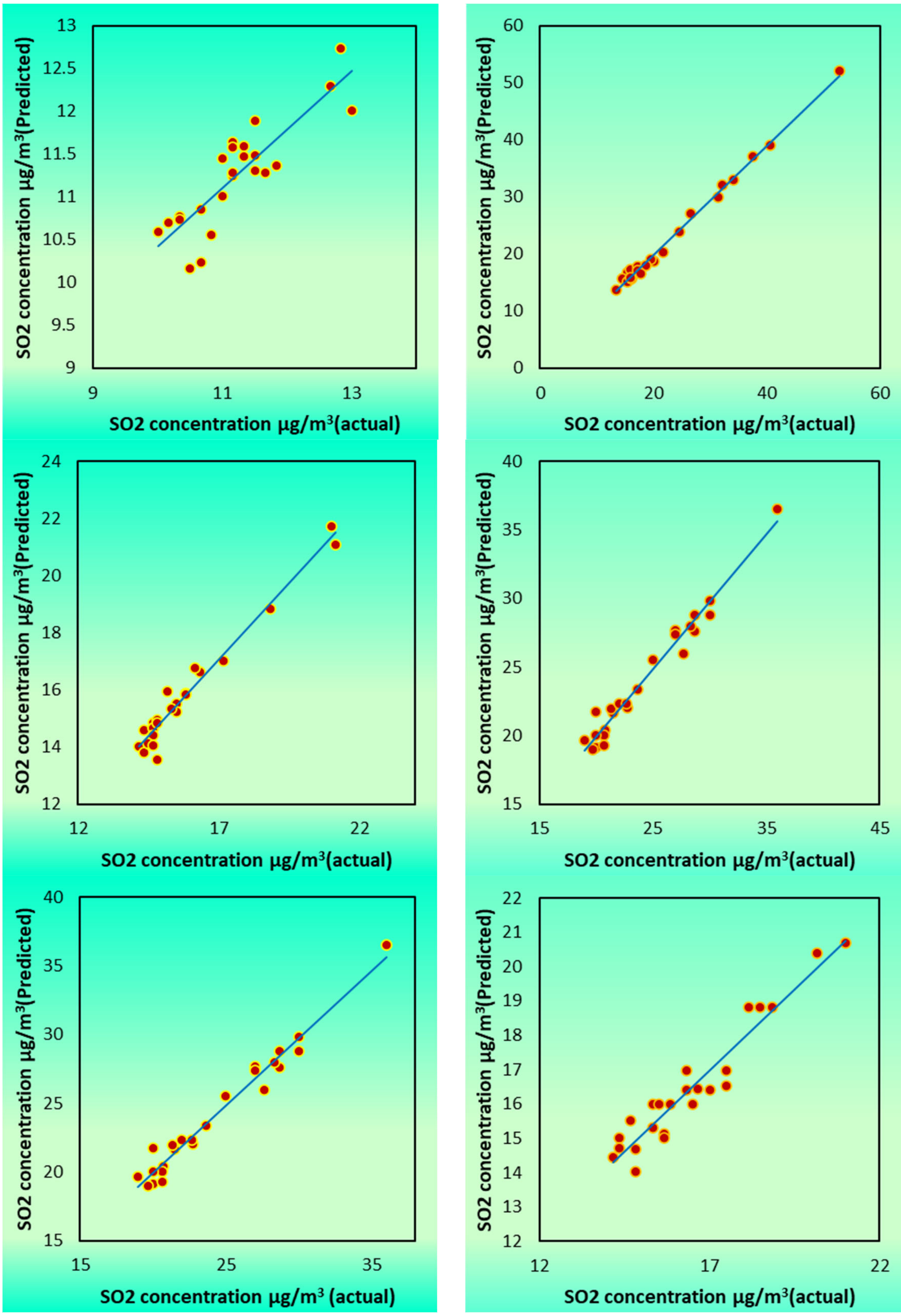


Fig. 5. Weekly average hourly concentration and prediction for summer (left) and winter (right) seasons for the year 2019, 2020 and 2021

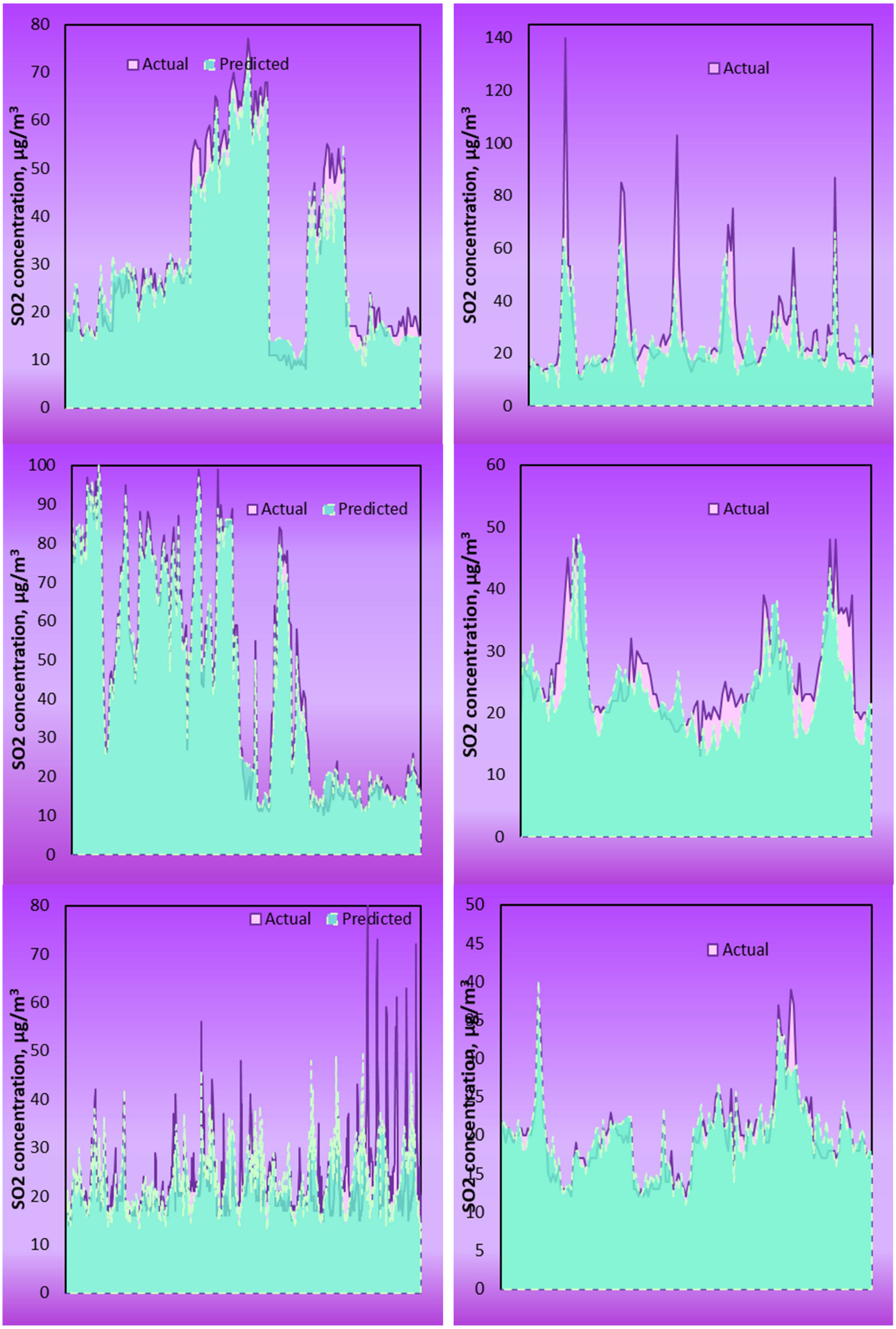


Fig. 6. Seasonal concentration of SO_2 left three for summer and right three for winters for years 2019, 2020 and 2021

4. Conclusion

This study evaluated the prediction performance of SO_2 for The town area of Mohammedia City, Morocco. LSTM neural network was used for prediction. The prediction was based on hourly concentration, weekly hourly concentration, and seasonal concentration in summer and winter for the years 2019, 2020, and 2021. The LSTM model can successfully memorise long-range temporal dependence (summer and winter) and short-range (hourly and weekly). Upon evaluating the model performance in terms of MAE, MAAPE, and RMSE, it was observed that the weekly hourly readings provided the most accurate results. Also, based on season, the LSTM model was more accurate for winter season SO_2 concentration than the summer season. The present study was limited owing to a lack of local incidence data. The study can aid policy and decision-makers in planning beforehand, considering the future scenario predicted by the LSTM model, and adopting mitigating measures accordingly.

Further studies are required to consider local incidences' impact on the prediction performance of the LSTM model as they cause sudden changes in the concentration of SO_2. Also, other artificial intelligence should be investigated to determine the most optimised model for evaluating the air quality of Mohammedia city in Morocco. Future studies must also Include other air quality parameters such as particulate matter, ozone, nitrogen dioxide, and carbon monoxide. Also, modelling needs to be done for air quality in Mohammedia City instead of just pollutant concentration.

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/40/45

Reference

Clare Heaviside, Claire Witham, Sotiris Vardoulakis (2020). Potential health impacts from sulphur dioxide and sulphate exposure in the UK resulting from an Icelandic effusive volcanic eruption. *Science of the Total Environment*, *774*, 140981. https://doi.org/10.1016/j.scitotenv.2021.145549

Haq, D.Z. et al. (2021). Long Short-Term Memory Algorithm for Rainfall Prediction Based on El-Nino and IOD Data. *Procedia Computer Science*, *179*(2019), 829-837. https://doi.org/10.1016/j.procs.2021.01.071

Li, X. et al. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, *231*, 997-1004. https://doi.org/10.1016/j.envpol.2017.08.114

Liu, Yu et al. (2021). State of charge prediction framework for lithium-ion batteries incorporating long short-term memory network and transfer learning. *Journal of Energy Storage*, *37*, 102494. https://doi.org/10.1016/j.est.2021.102494

Ma, J. et al. (2019). Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmospheric Environment*, *214*, 116885. https://doi.org/10.1016/j.atmosenv.2019.116885

Ma, J. et al. (2020). Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Science of the Total Environment*, *705*, 135771. https://doi.org/10.1016/j.scitotenv.2019.135771

Morakinyo, O.M., Mukhola, M.S., Mokgobu, M.I. (2020). Ambient gaseous pollutants in an urban area in South Africa: Levels and potential human health risk. *Atmosphere*, *11*(7). https://doi.org/10.3390/atmos11070751

Qi, Y. et al. (2019). A hybrid model for spatiotemporal forecasting of PM 2.5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, *664*, 1-10. https://doi.org/10.1016/j.scitotenv.2019.01.333

Seng, D. et al. (2021). Spatiotemporal prediction of air quality based on LSTM neural network. *Alexandria Engineering Journal*, *60*(2), 2021-2032. https://doi.org/10.1016/j.aej.2020.12.009

Wang, G. et al. (2021). A short-term voltage stability online prediction method based on graph convolutional networks and long short-term memory networks. *International Journal of Electrical Power and Energy Systems*, *127*, 106647. https://doi.org/10.1016/j.ijepes.2020.106647

Yang, G., Wang, Y., Li, X. (2020). Prediction of the NO_x emissions from thermal power plant using long-short term memory neural network. *Energy*, *192*, 116597. https://doi.org/10.1016/j.energy.2019.116597

Zhai, W., Cheng, C. (2020). A long short-term memory approach to predicting air quality based on social media data. *Atmospheric Environment*, *237*, 117411. https://doi.org/10.1016/j.atmosenv.2020.117411

Zhang, Luo et al. (2021). Air quality predictions with a semi-supervised bidirectional LSTM neural network. *Atmospheric Pollution Research*, *12*(1), 328-339. https://doi.org/10.1016/j.apr.2020.09.003

Zhang, Lei et al. (2021). Weather radar echo prediction method based on convolution neural network and Long Short-Term memory networks for sustainable e-agriculture. *Journal of Cleaner Production*, *298*, 126776. https://doi.org/10.1016/j.jclepro.2021.126776

Zhao, J. et al. (2019). Long short-term memory – Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. *Chemosphere*, *220*, 486-492. https://doi.org/10.1016/j.chemosphere.2018.12.128

Zuo, B., Cheng, J., Zhang, Z. (2021). Degradation prediction model for proton exchange membrane fuel cells based on long short-term memory neural network and Savitzky-Golay filter. *International Journal of Hydrogen Energy*, *46*(29), 15928-15937. https://doi.org/10.1016/j.ijhydene.2021.02.069